UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/600,797	06/20/2003	Eric D. Brill	MS303968.1 / MSFTP444US	9695

27195 7590 07/22/2008
AMIN. TUROCY & CALVIN, LLP
24TH FLOOR, NATIONAL CITY CENTER
1900 EAST NINTH STREET
CLEVELAND, OH 44114

EXAMINER	
HICKS, MICHAEL J	
ART UNIT	**PAPER NUMBER**
2165	

NOTIFICATION DATE	DELIVERY MODE
07/22/2008	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket1@thepatentattorneys.com
hholmes@thepatentattorneys.com
lpasterchek@thepatentattorneys.com

Office Action Summary	Application No.	Applicant(s)	
	10/600,797	BRILL, ERIC D.	
	Examiner	**Art Unit**	
	Michael J. Hicks	2165	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *03 April 2008*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-40,42 and 43* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-40,42 and 43* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *20 June 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1-40, 42 and 43 Pending.

Claim 41 Canceled.

Response to Arguments

2. Applicant's arguments, see response, filed 4/3/2008, with respect to the rejection(s) of claim(s) 1-40, 42 and 43 under USC 103(a) have been fully considered and are persuasive. Therefore, the rejection has been withdrawn. However, upon further consideration, a new ground(s) of rejection is made in view of the newly introduced art of Joachims ("Optimizing Search Engines Using Clickthrough Data", Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Pages 133-142, 2002, ACM) and the previously relied upon art of Pazzani.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

4. Claims 1-6, 8-16, 18-22, 29-40, and 42-43 rejected under 35 U.S.C. 102(a) as being anticipated by Joachims.

As per Claim 1, Joachims discloses a system that refines a general-purpose search engine, comprising: a component that identifies an entry point that includes a

link utilized to access the general-purpose search engine (i.e. *"To elicit data and provide a framework for testing the algorithm, I implemented a WWW meta-search engine called "Striver". Meta-search engines combine the results of several basic search engines without having a database of their own. Such a setup has several advantages. First, it is easy to implement while covering a large document collection - namely the whole WWW. Second, the basic search engines provide a basis for comparison."* The preceding text excerpt clearly indicates that an entry point including a link utilized to access at least one general purpose search engine (e.g. a metasearch engine) exists within the system.) (Page 137, Section 5.1); and a tuning component that receives search query results of the general-purpose search engine and filters the search results based at least on criteria associated with the entry point through which the general-purpose search engine was accessed (i.e. *"This paper presents an approach to learning retrieval functions by analyzing which links the users click on in the presented ranking. This leads to a problem of learning with preference examples like "for query q, document d, should be ranked higher than document db". More generally, I will formulate the problem of learning a ranking function over a finite domain in terms of empirical risk minimization. For this formulation, I will present a Support Vector Machine (SVM) algorithm that leads to a convex program and that can be extended to non-linear ranking functions. Experiments show that the method can successfully learn a highly effective retrieval function for a meta-search engine."* The preceding text excerpt clearly indicates that the results from the general purpose search engine are filtered based on ranking function (e.g. criteria associated with the entry point).) (Page 1, Introduction), the criteria comprises at least a first set of data categorized as relevant to a user's context and a second set of data categorized as non-relevant to the user's context (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-ID in the*

click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings." The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2), wherein user selection of a query result from a ranked list of the query results causes the selected result to be added to the first set of data and causes the results not selected by the user but ranked higher than the selected result to be automatically added to the second set of data (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently*

informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1), the first and second sets of data persisted to a computer-readable storage medium (i.e. *Examiner notes that as the training data and learned rating data accumulate over time (e.g. the system becomes more accurate over time), the first and second sets of data, used to determine relevancy, must be stored to a computer readable storage medium.*).

As per Claim 2, Joachims discloses the criteria comprising one or more of a document property, a context parameter, and a configuration (i.e. *"Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the criteria may comprise a document property (e.g. page rank or word occurrence), or context parameter (e.g. word probability).) (Page 136, Section 4.1).

As per Claim 3, Joachims discloses the document property comprising one or more of a term that appears on a web page, a property of a Uniform Resource Locator (URL) identifying the web page, a property of a plurality of URLs that link to the web page, a property of a plurality of web pages that link to the web page, and a layout (i.e. *"Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the criteria may comprise a document property (e.g. page rank or word occurrence), or context parameter (e.g. word probability).) (Page 136, Section 4.1).

As per Claim 4, Joachims discloses the context parameter comprising one of a word probability and a probability distribution (i.e. *"Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the criteria may comprise a document property (e.g. page rank or word occurrence), or context parameter (e.g. word probability).) (Page 136, Section 4.1).

As per Claim 5, Joachims discloses the tuning component is provided with training data to learn what properties of a document are indicative of the document being relevant to a user executing a search query from the entry point (i.e. *"This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the*

previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment" The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data.) (Page 138-139, Section 5.2).

As per Claim 6, Joachims discloses the tuning component configured to differentiate between a query result that is relevant to a search query context for a group of users and a query result that is non-relevant to the search query context for the group of users (i.e. *"Experimental results show that the algorithm performs well in practice, successfully adapting the retrieval function of a meta-search engine to the preferences of a group of users."* The preceding text excerpt clearly indicates that the system may be adapted to determine query relevance for a group of users.) (Page 141, Section 7).

As per Claim 8, Joachims discloses the tuning component generates one or more context parameters for a received query result, and compares the generated context parameters with a relevant context parameter and a non-relevant context parameter to determine whether the query result is relevant (i.e. *"Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the generated context parameters (e.g. word probability) are compared to context parameters in the relevant and non-relevant data sets.) (Page 136, Section 4.1).

As per Claim 9, Joachims discloses the tuning component further ranks the query results (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a*

document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search."* The preceding text excerpt clearly indicates that the tuning component ranks the query results.) (Page 135, Section 3).

As per Claim 10, Joachims discloses the ranking determined by the degree of relevance of the query result to the relevant data set and the non-relevant data set, the relevance is determined via one of a similarity measure and a confidence interval (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r* that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search… Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the ranking is determined by degree of relevance to the relevant and non-relevant data sets and that the relevance is determined, at least in part to similarity by a similarity measure.) (Page 135, Section 3; Page 136, Section 4.1).

As per Claim 11, Joachims discloses the ranking order comprising one of ascending and descending, from the most relevant result to the least relevant result (i.e. *Figure 3 clearly indicates that the results may be ranked in ascending order.*).

As per Claim 12, Joachims discloses the tuning component configured for a plurality of entry points associated with one or more groups of users (i.e. *"Experimental results show that the algorithm performs well in practice, successfully adapting the retrieval function of a meta-search engine to the preferences of a group of users… Furthermore, can clickthrough data also be used to adapt a search engine not to a group of users, but to the properties of a particular document collection? In particular, the factory-settings of any off-the-shelf retrieval system are necessarily suboptimal for any particular collection. Shipping off-the-shelf search engines with learning capabilities would enable them to optimize (and maintain) their performance automatically after being installed in a company intranet."* The preceding text excerpt clearly indicates that the tuning component may be configured to tune for particular entry points. Examiner notes that these entry points may be associated with specific groups of users or a specific user.) (Page 141, Section 7).

As per Claim 13, Joachims discloses a system that tunes a general-purpose search engine, comprising: a filter component that receives search query results of a general-purpose search engine and parses relevant and non-relevant results based on training data associated with the entry point that provides a link employed to traverse to the general-purpose search engine (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxv-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of*

Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings." The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2), the training data comprises a first set of data categorized as relevant to a search context of a user for the entry point and a second set of data categorized as non-relevant to the search context of the user (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This*

data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings." The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did not click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2), and a ranking component that sorts the filtered results in accordance with the training data for presentation to a user (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r* that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search."* The preceding text excerpt clearly indicates that the tuning component ranks the query results in accordance to the training data.) (Page 135, Section 3), wherein a user clicking a link associated with a search result from the sorted results causes the result to be added to the first set of data and causes the results whose links were not clicked by the user but that are ranked higher than the clicked result to be automatically added to the second set of data (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts*

presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1), the first and second sets of data persisted to a computer-readable storage medium (i.e. *Examiner notes that as the training data and learned rating data accumulate over time (e.g. the system becomes more accurate over time), the first and second sets of data, used to determine relevancy, must be stored to a computer readable storage medium.*).

As per Claim 14, Joachims discloses the filter component parses the results as a function of one or more of a document property, a context parameter, and a configuration associated with the entry point (i.e. *"Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly

indicates that the criteria may comprise a document property (e.g. page rank or word occurrence), or context parameter (e.g. word probability).) (Page 136, Section 4.1).

As per Claim 15, Joachims discloses the filter component trained to differentiate between a relevant and a non-relevant result via the training data (i.e. *"This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment"* The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data.) (Page 138-139, Section 5.2).

As per Claim 16, Joachims discloses the second set of data categorized as non-relevant comprising random data unrelated to the search context of the user for the entry point (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the auerv- ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance."* The preceding text excerpt clearly indicates that the unrelated data includes data relating to all queries the user has performed, or data from multiple queries in the training data set, and therefore includes random data unrelated to the search context of the user (e.g. the search results of unrelated queries).) (Page 134, Section 2.1).

As per Claim 18, Joachims discloses the ranking component employs a technique to determine the degree of relevance of the query results with respect to the relevant data set and the non-relevant data set (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r* that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search… Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the ranking is determined by degree of relevance to the relevant and non-relevant data sets and that the relevance is determined, at least in part to similarity by a similarity measure.) (Page 135, Section 3; Page 136, Section 4.1).

As per Claim 19, Joachims discloses the technique comprising one of a similarity measure and a confidence interval (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r* that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search… Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the ranking is determined by degree of relevance to the relevant and non-relevant data sets and that the relevance is determined, at least in part to similarity by a similarity measure.) (Page 135, Section 3; Page 136, Section 4.1).

As per Claim 20, Joachims discloses the ranking order comprising one of ascending and descending, from the most relevant result to the least relevant result (i.e. *Figure 3 clearly indicates that the results may be ranked in ascending order.*).

As per Claim 21, Joachims discloses the ranking performed on the relevant query results, the non-relevant results are discarded (i.e. *"The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user."* The preceding text excerpt clearly indicates that only relevant query results are ranked.) (Page 137, Section 5.1).

As per Claim 22, Joachims discloses a method to filter and rank general-purpose search engine results based on criteria associated with an entry point, comprising: executing a query search with the general-purpose search engine accessed through a link associated with the entry point (i.e. *"To elicit data and provide a framework for testing the algorithm, I implemented a WWW meta-search engine called "Striver". Meta-search engines combine the results of several basic search engines without having a database of their own. Such a setup has several advantages. First, it is easy to implement while covering a large document collection - namely the whole WWW. Second, the basic search engines provide a basis for comparison."* The preceding text excerpt clearly indicates that an entry point including a link utilized to access at least one general purpose search engine (e.g. a metasearch engine) exists within the system.) (Page 137, Section 5.1); filtering the general-purpose search engine results by tuning the general-purpose search engine based on a set of training data associated with the entry point employed to access the

general purpose search engine (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxv-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings."* The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2); and ranking the filtered general-purpose search engine results (i.e. *Figure 3 clearly indicates that the results may be ranked in ascending order.*).; automatically storing a first query result selected by a user in a first data set categorized as relevant (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7*

are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1); automatically storing at least one non-selected query result that is ranked higher than the first query result in a second data set categorized as non-relevant upon selection of the first query result (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that*

clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1); and including the first data set and second data set in the set of training data associated with the entry point employed to access the general purpose search engine (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are*

analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1).

As per Claim 29, Joachims discloses a method to customize a general-purpose search engine to improve context search query results, comprising: tuning a general-purpose search engine for an entry point by employing a method further comprising (i.e. *"To elicit data and provide a framework for testing the algorithm, I implemented a WWW meta-search engine called "Striver". Meta-search engines combine the results of several basic search engines without having a database of their own. Such a setup has several advantages. First, it is easy to implement while covering a large document collection - namely the whole WWW. Second, the basic search engines provide a basis for comparison. The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1."* The preceding text excerpt clearly indicates that an entry point including a link utilized to access at least one general purpose search engine (e.g. a metasearch engine) which is tuned exists within the system.) (Page 137, Section 5.1): providing a first set of data categorized as relevant that is used by a component to

discern query results relevant to a search context of a user employing the entry point (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings."* The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2), the entry point provides a link employed to access the general-purpose search engine (i.e. *"To elicit data and provide a framework for testing the algorithm, I implemented a WWW meta-search engine called "Striver". Meta-search engines combine the results of several basic*

search engines without having a database of their own. Such a setup has several advantages. First, it is easy to implement while covering a large document collection - namely the whole WWW. Second, the basic search engines provide a basis for comparison. The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that an entry point including a link utilized to access at least one general purpose search engine (e.g. a metasearch engine) which is tuned exists within the system.);

providing a second set of data categorized as non-relevant that is used by the component to discern query results unrelated to the search context (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating*

that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings." The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2), the first set of data and the second set of data are manually provided (i.e. *"This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001."* The preceding text excerpt clearly indicates that the training data may be manually provided.) (Page 138, Section 5.2); determining whether a query result is relevant or non-relevant to the search context based on the first set of relevant data and the second set of non-relevant data, each query result is compared with both the first set of data and second set of data to determine the relevance of the query result (i.e. *"Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the generated context parameters (e.g. word probability) are compared to context parameters in the relevant and non-relevant data sets.) (Page 136, Section 4.1); executing a search query with the general purpose search engine to obtain a ranked list of query results (i.e. *"The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents*

the top 50 links to the user." The preceding text excerpt clearly indicates that a query is executed with the general purpose search engine to return a ranked set of results.) (Page 137, Section 5.1); selecting a link associated with a query result from the list (i.e. *"The clicks of the user are recorded using the proxy system described in Section 2.1."* The preceding text excerpt clearly indicates that a link associated with a query result is selected from the list.) (Page 137, Section 5.1); automatically adding the selected query result to the first set of data (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1."* The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1); and automatically adding non-selected results from the list that are ranked higher than the selected query result to the second set of data upon selection of

the selected query result (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1."* The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1).

As per Claim 30, Joachims discloses the first set of data categorized as relevant comprising data associated with the search context of the user for the entry point (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-*

ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance… Experimental results show that the algorithm performs well in practice, successfully adapting the retrieval function of a meta-search engine to the preferences of a group of users" The preceding text excerpt clearly indicates that the data categorized as relevant is associated with the search context of the user, or group of users.) (Page 134, Section 2.1; Page 141, Section 7).

As per Claim 31, Joachims discloses the second set data categorized as non-relevant comprising random data unrelated to the search context of the user for the entry point (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxv-server records the URL and the auerv- ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance."* The preceding text excerpt clearly indicates that the unrelated data includes data relating to all queries the user has performed, or data from multiple queries in the training data set, and therefore includes random data unrelated to the search context of the user (e.g. the search results of unrelated queries).) (Page 134, Section 2.1).

As per Claim 32, Joachims discloses providing information to associate respective query results with the entry point (i.e. *"While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search."* The preceding text excerpt clearly indicates that information may be provided to associate the query results with a certain entry point or user. Examiner notes that the query log and

clickthrough log also associate the query results to the entry point, as it is stored at the entry point.) (Page 135, Section 3).

As per Claim 33, Joachims discloses the first set of data categorized as relevant and the second set of data categorized as non-relevant employed to train the component to learn the features that differentiate relevant data from non-relevant data (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r* that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search… Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the ranking is determined by degree of relevance to the relevant and non-relevant data sets and that the relevance is determined, at least in part to similarity by a similarity measure.) (Page 135, Section 3; Page 136, Section 4.1).

As per Claim 34, Joachims discloses a method to automatically customize a general-purpose search engine for an entry point, comprising: identifying the entry point (i.e. *"To elicit data and provide a framework for testing the algorithm, I implemented a WWW meta-search engine called "Striver". Meta-search engines combine the results of several basic search engines without having a database of their own. Such a setup has several advantages. First, it is easy to implement while covering a large document collection - namely the whole WWW. Second, the basic search engines provide a basis for comparison. The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" ,*

and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that an entry point including a link utilized to access at least one general purpose search engine (e.g. a metasearch engine) which is tuned exists within the system.) (Page 137, Section 5.1); executing a query search via the entry point that includes a link employed to route to the general-purpose search engine (i.e. *"To elicit data and provide a framework for testing the algorithm, I implemented a WWW meta-search engine called "Striver". Meta-search engines combine the results of several basic search engines without having a database of their own. Such a setup has several advantages. First, it is easy to implement while covering a large document collection - namely the whole WWW. Second, the basic search engines provide a basis for comparison. The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1."* The preceding text excerpt clearly indicates that an entry point including a link utilized to access at least one general purpose search engine (e.g. a metasearch engine) which is tuned exists within the system.) (Page 137, Section 5.1); recording a first query result from a ranked list of query results returned from the executed query selected by a user employing the entry point as relevant when a user views the document associated with the first query result (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2*

with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch", "Excite", "Altavista", and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1); recording at least one second query result whose associated document was not viewed by the user but that is ranked higher than the first query result as non-relevant when ranked the first result is selected for viewing by the user (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance*

judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch", "Excite", "Altavista", and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1); and providing the recorded results to automatically train the filter for the entry point, in order to discriminate between results relevant to a search context of the user for the entry point and results non-relevant to the search context (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by*

these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1).

As per Claim 35, Joachims discloses the set of relevant data comprising data associated with the search context of the user for the entry point (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance… Experimental results show that the algorithm performs well in practice, successfully adapting the retrieval function of a meta-search engine to the preferences of a group of users"* The preceding text excerpt clearly indicates that the data categorized as relevant is associated with the search context of the user, or group of users.) (Page 134, Section 2.1; Page 141, Section 7).

As per Claim 36, Joachims discloses the set of non-relevant data comprising data unrelated to the search context of the user for the entry point (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented*

ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance." The preceding text excerpt clearly indicates that the unrelated data includes data relating to all queries the user has performed, or data from multiple queries in the training data set, and therefore includes random data unrelated to the search context of the user (e.g. the search results of unrelated queries).) (Page 134, Section 2.1).

As per Claim 37, Joachims discloses providing information to associate respective query results with the entry point (i.e. *"While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search."* The preceding text excerpt clearly indicates that information may be provided to associate the query results with a certain entry point or user. Examiner notes that the query log and clickthrough log also associate the query results to the entry point, as it is stored at the entry point.) (Page 135, Section 3).

As per Claim 38, Joachims discloses the set of relevant data and the set of non-relevant data employed to train the component to learn the features that differentiate relevant data from non-relevant data (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r* that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search… Such features are, for example, the*

number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)." The preceding text excerpt clearly indicates that the ranking is determined by degree of relevance to the relevant and non-relevant data sets and that the relevance is determined, at least in part to similarity by a similarity measure.) (Page 135, Section 3; Page 136, Section 4.1).

As per Claim 39, Joachims discloses the query results selected via a click thru technique employing a mouse to select a link associated with the query result by clicking on the link (i.e. *"When the user clicks on the link, the proxy-server records the URL and the query-ID in the click-log."* The preceding text excerpt clearly indicates that the query results are selected by employing a mouse to click on a link associated with the query result.) (Page 134, Section 2.1).

As per Claim 40, Joachims discloses generating a word probability distribution for the relevant recorded results and a word probability distribution for the non-relevant recorded results (i.e. *"Such features are, for example, the number of words that query and document share, the number of words they share inside certain HTML tags (e.g. TITLE, HI, H2, ...), or the page-rank of d [22] (see also Section 5.2)."* The preceding text excerpt clearly indicates that the criteria used to evaluate the relevant and non-relevant data may comprise a document property (e.g. page rank or word occurrence), or context parameter (e.g. word probability).) (Page 136, Section 4.1).

As per Claim 42, Joachims discloses a computer readable storage medium storing computer executable components that tunes a general-purpose search engine to improve context search query results, comprising: a component that receives search query results of a general-purpose search engine and filters the results based on

training data sets associated with the search context of a user depending on the entry point that provides a link utilized to arrive at the general-purpose search engine (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings."* The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2), the training data sets include at least a first category of data explicitly defined to be relevant to the search context and a second category of data explicitly

defined to be non-relevant to the search context (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxv-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings."* The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2); and a component that ranks the filtered general-purpose search engine results according to the similarity of the search engine results to the training data sets (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r**

that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search." The preceding text excerpt clearly indicates that the tuning component ranks the query results in accordance to the training data.) (Page 135, Section 3), wherein selecting a link associated with a first search result from the ranked results causes the first result to be added to the first set of data and causes results that are ranked higher than the first result and have not been selected by the user to be automatically added to the second set of data (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6…The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1."* The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1).

As per Claim 43, Joachims discloses a system that receives, filters and ranks general-purpose search engine results, comprising: means for filtering general-purpose search engine results by determining whether a query result is relevant to a search context of a group of users, the search context is associated with an entry point that includes a link employed to navigate to the general-purpose search engine (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxv-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings."* The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that

the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page 134, Section 2.1; Page 138-139, Section 5.2), the search context further having an associated first set of training data categorized as relevant to the context and an associated second set of training data categorized as non-relevant to the context (i.e. *"Each query is assigned a unique ID which is stored in the query-log along with the query words and the presented ranking. The links on the results-page presented to the user do not lead directly to the suggested document, but point to a proxy server. These links encode the query-ID and the URL of the suggested document. When the user clicks on the link, the proxy-server records the URL and the query-ID in the click-log. The proxy then uses the HTTP-Location command to forward the user to the target URL. This process can be made transparent to the user and does not influence system performance…This experiment verifies that the Ranking SVM can indeed learn regularities using partial feedback from clickthrough data. To generate a first training set, I used the Striver search engine for all of my own queries during October, 2001. Striver displayed the results of Google and MSNSearch using the combination method from the previous section. All clickthrough triplets were recorded. This resulted in 112 queries with a non-empty set of clicks. This data provides the basis for the following offline experiment...From the 112 queries, pairwise preferences were extracted according to Algorithm 1 described in Section 2.2. In addition, 50 constraints were added for each clicked-on document indicating that it should be ranked higher than a random other document in the candidate set V. While the latter constraints are not based on user feedback, they should hold for the optimal ranking in most cases. These additional constraints help stabilize the learning result and keep the learned ranking function somewhat close to the original rankings."* The preceding text excerpt clearly indicates that the criteria comprises at least a set of relevant (e.g. as defined by a training data set, or gathered during system operation) and non-relevant data. Examiner notes that the relevant data is the data clicked on by the user while the non-relevant data is the data which the user did no click on. Examiner further notes that the non-selected data may be determined to be related to the search query by the metasearch engines initial retrieval, but relevance is determined by user clickthrough data.) (Page

134, Section 2.1; Page 138-139, Section 5.2); and means for ranking the filtered general-purpose search engine results based on a relevance of the general-purpose search engine results to the search context of the group of users and the entry point as determined by a comparison of the search engine results with the first and second sets of training data (i.e. *"The problem of information retrieval can be formalized as follows. For a query q and a document collection D = {dl, ..., dm), the optimal retrieval system should return a ranking r* that orders the documents in D according to their relevance to the query. While the query is often represented as merely a set of keywords, more abstractly it can also incorporate information about the user and the state of the information search."* The preceding text excerpt clearly indicates that the tuning component ranks the query results in accordance to the training data.) (Page 135, Section 3), wherein a user viewing a document associated with a first search result from the ranked results causes the first result to be added to the first set of training data and causes the results that are unviewed but ranked higher than the first result to be automatically added to the second set of training data (i.e. *"Consider again the example from Figure 1. While it is not possible to infer that the links 1, 3, and 7 are relevant on an absolute scale, it is much more plausible to infer that link 3 is more relevant than link 2 with probability higher than random. Assuming that the user scanned the ranking from top to bottom, he must have observed link 2 before clicking on 3, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently informative, this gives some indication of the user's preferences. Similarly, it is possible to infer that link 7 is more relevant than links 2, 4, 5, and 6. This means that clickthrough data does not convey absolute relevance judgments, but partial relative relevance judgments for the links the user browsed through. A search engine ranking the returned links according to their relevance to q should have ranked links 3 ahead of 2, and link 7 ahead of 2, 4, 5, and 6...The "Striver" meta-search engine works as follows. The user types a query into Striver's interface. This query is forwarded to "Google", "MSNSearch" , "Excite", "Altavista" , and "Hotbot". The results pages returned by these basic search engines are analyzed and the top 100 suggested links are extracted. After*

canonicalizing URLs, the union of these links composes the candidate set V. Striver ranks the links in V according to its learned retrieval function faw and presents the top 50 links to the user. For each link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system described in Section 2.1." The preceding text excerpt clearly indicates that selected results from the candidate set V are recorded as relevant and non-selected results, including those ranked higher than the selected results) are recorded as non-relevant (e.g. the set of V, not selected by the user, but written to the query log.).) (Page 135, Section 2.2; Page 137, Section 5.1), the first and second sets of training data stored on a computer-readable storage medium (i.e. *Examiner notes that as the training data and learned rating data accumulate over time (e.g. the system becomes more accurate over time), the first and second sets of data, used to determine relevancy, must be stored to a computer readable storage medium.*).

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. Claims 7, 17, and 23-28 rejected under 35 U.S.C. 103(a) as being unpatentable over Joachims in view of Pazzani.

As per Claim 7, Joachims fails to disclose the tuning component employs statistical analysis in connection with filtering the search query results.

Pazzani discloses the tuning component employs statistical analysis in connection with filtering the search query results *(i.e. Page 319, Paragraph 2 indicates that statistical analysis (e.g. probability calculations) are employed in connection with the filtering.)*.

It would have been obvious to one skilled in the art at the time of Applicants invention to modify the teachings of Joachims with the teachings of Pazzani to include the tuning component employs statistical analysis in connection with filtering the search query results with the motivation of learning and revising user profiles that can determine which World Wide Web sites on a given topic would be interesting to a user (Pazzani, Abstract).

As per Claim 17, Joachims fails to disclose the filter component employs statistical analysis to determine whether a result is relevant or non-relevant to the entry point

Pazzani discloses the filter component employs statistical analysis to determine whether a result is relevant or non-relevant to the entry point *(i.e. Page 319, Paragraph 2 indicates that statistical analysis (e.g. probability calculations) are employed in connection with the filtering.)*.

It would have been obvious to one skilled in the art at the time of Applicants invention to modify the teachings of Joachims with the teachings of Pazzani to include the filter component employs statistical analysis to determine whether a result is relevant or non-relevant to the entry point with the motivation of learning and revising

user profiles that can determine which World Wide Web sites on a given topic would be interesting to a user (Pazzani, Abstract).

As per Claim 23, Joachims fails to disclose employing a statistical hypothesis to determine whether a result is relevant or non-relevant to a search context of the entry point.

Pazzani discloses employing a statistical hypothesis to determine whether a result is relevant or non-relevant to a search context of the entry point *(See Page 317, Paragraph 2 which indicates that a statistical hypothesis (e.g. conversion to positive and negative feature vectors) is used to determine whether a result is relevant or non-relevant.)*.

It would have been obvious to one skilled in the art at the time of Applicants invention to modify the teachings of Joachims with the teachings of Pazzani to include employing a statistical hypothesis to determine whether a result is relevant or non-relevant to a search context of the entry point with the motivation of learning and revising user profiles that can determine which World Wide Web sites on a given topic would be interesting to a user (Pazzani, Abstract).

As per Claim 24, Joachims fails to disclose the statistical hypothesis employing a threshold in connection with a probability distribution for relevant data and a probability distribution for non-relevant data, respective word probabilities are generated for the search query results and compared to the threshold, the probability distribution for

relevant data and the probability distribution for non-relevant data to determine whether the results are relevant or non-relevant.

Pazzani discloses the statistical hypothesis employing a threshold in connection with a probability distribution for relevant data and a probability distribution for non-relevant data, respective word probabilities are generated for the search query results and compared to the threshold, the probability distribution for relevant data and the probability distribution for non-relevant data to determine whether the results are relevant or non-relevant (*See page 319, Paragraph 2, which indicates that a statistical probability hypothesis is employed to determine relevance. Note that there must exist some threshold which indicates the separation between relevance and non-relevance.*).

It would have been obvious to one skilled in the art at the time of Applicants invention to modify the teachings of Joachims with the teachings of Pazzani to include the statistical hypothesis employing a threshold in connection with a probability distribution for relevant data and a probability distribution for non-relevant data, respective word probabilities are generated for the search query results and compared to the threshold, the probability distribution for relevant data and the probability distribution for non-relevant data to determine whether the results are relevant or non-relevant with the motivation of learning and revising user profiles that can determine which World Wide Web sites on a given topic would be interesting to a user (Pazzani, Abstract).

As per Claim 25, Joachims fails to disclose the threshold employed to bias the decision to mitigate one of a result being deemed non-relevant when the result is relevant and a result being deemed relevant when the result is non-relevant.

Pazzani discloses the threshold employed to bias the decision to mitigate one of a result being deemed non-relevant when the result is relevant and a result being deemed relevant when the result is non-relevant (*See page 319, Paragraph 2, which indicates that a statistical probability hypothesis is employed to determine relevance. Note that there must exist some threshold which indicates the separation between relevance and non-relevance.*).

It would have been obvious to one skilled in the art at the time of Applicants invention to modify the teachings of Joachims with the teachings of Pazzani to include the threshold employed to bias the decision to mitigate one of a result being deemed non-relevant when the result is relevant and a result being deemed relevant when the result is non-relevant with the motivation of learning and revising user profiles that can determine which World Wide Web sites on a given topic would be interesting to a user (Pazzani, Abstract).

As per Claim 26, Joachims fails to disclose further employing a probability distribution analysis or machine learning in connection with the filtering and ranking, wherein suitable probability distributions include a Bernoulli, a binomial, a Pascal, a Poisson, an arcsine, a beta, a Cauchy, a chi-square with N degrees of freedom, an Erlang, a uniform, an exponential, a gamma, a Gaussian-univariate, a Gaussian-bivariate, a Laplace, a log-normal, a rice, a Weibull and a Rayleigh distribution, and the

machine learning can classify based on one or more of a word occurrence, a distribution, a page layout, an inlink, and an outlink.

Pazzani discloses further employing a probability distribution analysis or machine learning in connection with the filtering and ranking, wherein suitable probability distributions include a Bernoulli, a binomial, a Pascal, a Poisson, an arcsine, a beta, a Cauchy, a chi-square with N degrees of freedom, an Erlang, a uniform, an exponential, a gamma, a Gaussian-univariate, a Gaussian-bivariate, a Laplace, a log-normal, a rice, a Weibull and a Rayleigh distribution (*See Page 319, Paragraph 2, which indicates the use of a uniform probability distribution.*), and the machine learning can classify based on one or more of a word occurrence, a distribution, a page layout, an inlink, and an outlink (*See Page 317, Paragraphs 2-4 Page which indicate the use of word occurrence.*).

It would have been obvious to one skilled in the art at the time of Applicants invention to modify the teachings of Joachims with the teachings of Pazzani to include further employing a probability distribution analysis or machine learning in connection with the filtering and ranking, wherein suitable probability distributions include a Bernoulli, a binomial, a Pascal, a Poisson, an arcsine, a beta, a Cauchy, a chi-square with N degrees of freedom, an Erlang, a uniform, an exponential, a gamma, a Gaussian-univariate, a Gaussian-bivariate, a Laplace, a log-normal, a rice, a Weibull and a Rayleigh distribution, and the machine learning can classify based on one or more of a word occurrence, a distribution, a page layout, an inlink, and an outlink with the motivation of learning and revising user profiles that can determine which World Wide Web sites on a given topic would be interesting to a user (Pazzani, Abstract).

As per Claim 27, Joachims fails to disclose employing a statistical analysis to rank search query results.

Pazzani discloses employing a statistical analysis to rank search query results (i.e. *Page 319, Paragraph 2 which indicates that the classifier can be used to rank order pages by returning a probability (e.g. a statistical analysis).*).

It would have been obvious to one skilled in the art at the time of Applicants invention to modify the teachings of Joachims with the teachings of Pazzani to include employing a statistical analysis to rank search query results with the motivation of learning and revising user profiles that can determine which World Wide Web sites on a given topic would be interesting to a user (Pazzani, Abstract).

As per Claim 28, Joachims fails to disclose the ranking comprising one of generating word probabilities and employing a confidence interval to determine relevance, and generating a similarity measure comprising one of a cosine distance, the Jaccard coefficient, an entropy-based measure, a divergence measure and/or a relative separation measure to determine similarity.

Pazzani discloses the ranking comprising one of generating word probabilities and employing a confidence interval to determine relevance (*See Page 316, Paragraphs 2-3 which indicate the use of a confidence interval to determine applicable words and word probabilities.*), and generating a similarity measure comprising one of a cosine distance, the Jaccard coefficient, an entropy-based measure, a divergence measure and/or a relative

separation measure to determine similarity (*See Page 319 which indicates the use of a separation measure (e.g. probability scale) in the ranking.*).

It would have been obvious to one skilled in the art at the time of Applicants invention to modify the teachings of Joachims with the teachings of Pazzani to include the ranking comprising one of generating word probabilities and employing a confidence interval to determine relevance, and generating a similarity measure comprising one of a cosine distance, the Jaccard coefficient, an entropy-based measure, a divergence measure and/or a relative separation measure to determine similarity with the motivation of learning and revising user profiles that can determine which World Wide Web sites on a given topic would be interesting to a user (Pazzani, Abstract).

7. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Points of Contact

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael J. Hicks whose telephone number is (571) 272-2670. The examiner can normally be reached on Monday - Friday 9:00a - 5:30p.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Christian Chace can be reached on (571) 272-4190. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Michael J Hicks
Art Unit 2165
Phone: (571) 272-2670
Fax: (571) 273-2670

/Christian P. Chace/

Supervisory Patent Examiner, Art Unit 2165